Genetic variants associated with Crohn’s disease

Sonia Michail¹
Gilberto Bultron¹
R William DePaolo²

¹The University of Southern California, Children’s Hospital of Los Angeles, Los Angeles, CA, USA; ²Molecular Microbiology and Immunology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Abstract: Crohn’s disease is an immune-related disorder characterized by inflammation of the gastrointestinal mucosa, which can occur in any area throughout the digestive tract. This life-long disease commonly presents with abdominal pain, diarrhea, vomiting, and weight loss. While the exact etiology of this disease is largely unknown, it is thought to arise from an interaction between microbial, immunological, and environmental factors in a genetically susceptible host, whereby the immune system attacks the intestine as it cross reacts against gut microbial antigens. The study of genetic variants associated with Crohn’s disease has shed light on our understanding of disease pathophysiology. A large number of genetic variants identified in Crohn’s disease are related to genes targeting microbial recognition and bacterial wall sensing, the most common being NOD2/CARD15 gene. This review will discuss the recent advance in our knowledge of genetic variants of this disease and how they influence the disease course and prognosis.

Keywords: Crohn’s disease, genetics, autophagy

Introduction

Inflammatory bowel disease (IBD) is defined as a group of chronic intestinal conditions characterized by inflammation in the intestine. The most common types of IBD are Crohn’s disease (CD) and ulcerative colitis (UC). CD can occur in any area along the alimentary tract, but UC is localized primarily to the large intestine. Full thickness of the bowel wall is often affected in CD, while only the mucosa is involved in UC. IBD is a common disorder worldwide, with a prevalence of 0.1%–0.4% of the general population.¹ In the United States, the Center for Disease Control estimates that over 1.4 million Americans are affected by this disease, with an overall estimated annual health care cost exceeding US$1.7 billion (cdc.gov/ibd/). IBD is currently without a cure, requires life-long care, interferes with quality of life, and frequently necessitates surgical intervention. Therefore, it has been recognized as one of the top five gastrointestinal disease burdens in the USA. Recent international trends suggest that there is an increase in the incidence of CD.² Whether this finding represents a true rise in the incidence of the disease or merely reflects an improvement in diagnostic and reporting methods is difficult to determine at the present time. The past few decades have heralded new discoveries in IBD, most of which are related to the genetics of the disease, diagnostic tools, and management strategies. Recent discoveries have brought much attention to the genetic predisposition of patients with IBD. Those genetic variations are largely related to the innate immune system and microbial recognition.³–⁷
Genetic variants in Crohn’s disease

CD is a complex, multifactorial disease with strong genetic associations that lead to aberrant, chronic mucosal inflammation after exposure to environmental or microbial triggers. Thus far, 163 risk loci have been identified in patients with IBD, many of which are involved in innate immune responses and mucosal homeostasis. Currently, our genetic knowledge of CD is applied to understanding the natural history of the disease, describing racial predominance, predicting early onset strictureing and fistulizing, and determining location and overall severity of disease. Wagner et al recently described an association between mutations in toll-like receptor (TLR) 4 (rs4986790) and interleukin 10 receptor, alpha subunit (rs22291130) with Mycobacterium avium subspecies paratuberculosis positive CD patients, further implicating the role of microbial triggers and aberrant genetic responses in the immune system. This genetic data is helpful information that will possibly lead to the ultimate discovery of etiologies for the disease.

Significant CD risk has been associated with NOD2/CARD15, IBD5, and DLG5; IL23R and ATG16L1 have also been identified as CD susceptibility genes (see Table 1 for gene expansions). Genetic variants of Immunity-related GTPase family M (IRGM) have been significantly associated with CD, and it will be discussed later in this review, along with the other variant in the autophagy pathway, ATG16L1.

Three single nucleotide polymorphisms (SNPs) found within the NOD2/CARD15 gene (R702W, G908R, and 1007fsinsC) were found to be racially distinct. These SNPs have been established as independent risk factors for CD in Caucasians. Biank et al reviewed ten manuscripts, which together included 1319 children with NOD2/CARD15 mutations. They concluded that Caucasian children with European ancestry have similar attributable risks for developing CD as adults due to NOD2/CARD15. Furthermore, as stated previously, and currently restated by Biank et al, the NOD2/CARD15 mutational variants are nearly absent or rare in East Asians, Arabs, Africans, and African-Americans.

Genome wide association studies (GWAS) have been used very effectively in IBD genetic research. Deep resequencing of GWAS data has identified highly significant variants at CARD9, NOD2/CARD15, CUL2, and IL18RAP, which contribute to risk associations of previously defined variants at these loci, and as well, showed the functionality of the newly implicated NOD2/CARD15 variants. Rivas et al went on to describe additional protective variants at IL23R, and Stoll et al showed DLG5 R30Q is a modest risk gene for CD.

Further analysis of the data by Stoll et al revealed gender differences in allele frequency of R30Q between adult male CD cases and controls. They concluded that the DLG5 R30Q variant confers increased risk to adult males for CD. Biank et al compared the results of their pediatric CD study sample to the adult CD study sample results reported by Friedrichs et al and found the allele frequency of R30Q in children with CD to be 10.7% for males and 5.6% for females. The allele frequency of R30Q for controls was 8.3% for males and 12.3% for females. This difference of allele frequencies among females with CD and controls shows that R30Q may have a protective effect in female children.

Innate cytokines produced by antigen presenting cells in response to microbial stimulation are critical for the induction of specific T effector responses that can combat invading bacteria or viruses. Interleukin (IL)-12 is an innate cytokine composed of a p40 and a p35 subunits, which are essential for the polarization of interferon-γ-producing T helper 1 (Th1) cells. Interestingly, a second innate cytokine called IL-23 was discovered in 2000, and it shares the p40 subunit of IL-12, but instead of p35, it heterodimerizes with a unique p19 chain. Early studies examining IL-12 deletion of the p40 chain or via neutralization by antibodies targeting the p40 chain may have also effected IL-23-mediated responses. However, over the last 12 years IL-23 has emerged as a proinflammatory driver in gastrointestinal inflammation. IL-12 is important for Th1 and interferon-γ responses, whereas IL-23 has been shown to expand IL-17 producing Th17 cells and IL-17-producing innate lymphoid cells via expression of IL-23R.

The IL-23/IL-17 inflammatory axis has now been implicated in the pathogenesis of Crohn’s disease. This conclusion comes in part from the findings that CD patients have a higher number of circulating IL-17+ cells and from a study by Rismo et al, which showed that increased messenger (m) RNA expression of tumor necrosis factor (TNF) and IL17A in healed mucosa significantly increased the risk of relapse (hazard ratio [HR] = 3.4, P = 0.03, sensitivity 80%, specificity 38% and HR = 4.1, P = 0.008, sensitivity 81%, specificity 61%, respectively).

More thorough investigations linking genetic variants of IL23, IL17, NOD2/CARD15, IRGM, ATG16L1, and DLG5 to mRNA expression, protein expression, gender, age, race, and phenotypic outcomes will help us to further understand the etiologies of Crohn’s disease. A comprehensive list of the main genetic variants in CD is summarized in Table 1. In summary, the mutational variants in CD further describe a defect in bacterial sensing, as mentioned by Wagner et al, and an exaggerated response of the intestinal immune system.
Table 1 Most clinically relevant and GWAS replicated CD genes

<table>
<thead>
<tr>
<th>GENE</th>
<th>SNP or mutational variant</th>
<th>Function</th>
<th>Phenotype/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLR4</td>
<td>Asp299Gly</td>
<td>Lymphocyte lineage</td>
<td></td>
</tr>
<tr>
<td>IL23R</td>
<td>RS100481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCTN1</td>
<td>rs1050152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATG16L1</td>
<td>T300A (Asp299Gly)</td>
<td>Mucosal homeostasis and cellular component degradation</td>
<td></td>
</tr>
<tr>
<td>IRGM</td>
<td>RS1174186</td>
<td>Regulation of autophagy formation in response to intracellular pathogens</td>
<td></td>
</tr>
<tr>
<td>DLG5</td>
<td>R30Q</td>
<td>Common in pediatric CD; Inversely related in female children</td>
<td></td>
</tr>
<tr>
<td>LRRK2</td>
<td>RS11175593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBDS</td>
<td>503F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTPN22</td>
<td>Rs247660</td>
<td>Responsiveness of T and B cell receptors</td>
<td>620W gain-of-function variant is associated with reduced risk of CD</td>
</tr>
<tr>
<td>IL10R</td>
<td>1082A (G65C)</td>
<td></td>
<td>Increased risk of CD</td>
</tr>
<tr>
<td>TNF alpha</td>
<td>308A (A702V)</td>
<td>Extra-intestinal associations</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Bolded genes are highly significant genes in multiple GWAS. Nucleotide-binding oligomerization domain-containing protein 2 (NOD2) also known as caspase recruitment domain-containing protein 15 (CARD15) or inflammatory bowel disease protein 1 (IBD1) is a protein that is involved in the inflammasome and plays an important role in recognizing bacterial molecules and stimulating an immune response. NOD2 is an intracellular pattern recognition receptor, which recognizes molecules containing the specific structure called muramyl dipeptide found in bacteria. TLR 4 is a Toll-like receptor. It detects lipopolysaccharide from gram-negative bacteria and is thus important in the activation of the innate immune system. Interleukin 23 receptor (IL23R), is a type I cytokine receptor. OCTN1 encoded protein is an integral protein of the plasma membrane responsible for the cotransport of sodium ions and ergothioneine, which is an antioxidant, into cells. Autophagy-related protein 16-1 (ATG16L1), is a protein that is essential for autophagy. Immunity-related GTPase family M protein (IRGM), which plays a role in the innate immune response by regulating autophagy formation in response to intracellular pathogens. Disks large homolog 5 (DLG5), which functions as a scaffolding molecule at sites of cell-cell contact. The protein encoded by this gene localizes to the plasma membrane and cytoplasm and interacts with components of adherens junctions and the cytoskeleton. It is proposed to function in the transmission of extracellular signals to the cytoskeleton and in the maintenance of epithelial cell structure. Leucine-rich repeat kinase 2 (LRRK2), encodes a protein which is a substrate for chaperone-mediated autophagy. The RS11175593 SNP for LRRK2 just downstream MUC 19 encodes for protective mucus on epithelium. IBDS is a 250 kb cytokine gene cluster on chromosome 5q31 and associated with risk of developing Crohn’s disease. Protein tyrosine phosphatase, non-receptor type 2 (PTPN22) affects the responsiveness of T and B cell receptors, and mutations are associated with autoimmune diseases. Interleukin 1 receptor (IL10R), the protein encoded by this gene is a receptor for interleukin 10, which has been shown to mediate the immunosuppressive signal of interleukin 10, and thus inhibits the synthesis of proinflammatory cytokines. Tumor necrosis factor-alpha (TNF-alpha) is a cytokine involved in systemic inflammation.

**Abbreviations:** CD, Crohn’s disease; GWAS, genome-wide association studies; SNP, single nucleotide polymorphism; TLR, toll-like receptor; OCTN1, organic cation transporter novel type 1; GTP, guanosine triphosphate; MUC, mucin gene; IBD, inflammatory bowel disease

Autophagy in Crohn’s disease

Autophagy is a conserved cellular pathway that maintains cellular homeostasis via the degradation of cytosolic contents and proteins. This process occurs via the sequestration of cytoplasmic contents into autophagosomes and the subsequent fusion with lysosomes for degradation and recycling (Figure 1). Autophagy has drawn interest beyond the extent of homeostatic regulation and is now a widely recognized pathophysiological process in autoimmune diseases, and in cancer, and IBD. Intestinal tissues are in close proximity to the microbial world and therefore rely on intracellular defense mechanisms like autophagy in order to prevent aberrant or harmful inflammation. Thus, defects in directing commensals and pathogens to the autophagy pathway may impair bacterial responses and promote an environment in which antimicrobial immunity and inflammation is favored.

ATG16L1 gene variant rs2241880 (T300A) is a non-synonomous coding SNP that results in a single amino acid change in the mature protein. rs2241880 was identified in both a transmission disequilibrium test and a case controlled comparison in a study by Hampe et al examining CD patients in a German cohort. These data suggest that the CD risk conferred by ATG16L1 gene variation is confined to individuals carrying allele G, with a 60% frequency of the G allele in affected individuals compared to 53% in controls. These data were confirmed in patients from the United Kingdom (G allele frequency was 59% affected versus 52% in controls). Interestingly, a statistical interaction between rs2241880 and the CARD15 genotypes was discovered. This genetic association is further supported by the findings that muramyl dipeptide enhances autophagy and is inhibited by ATG16L1CD risk allele. Additionally, the ATG16L1 (T300A) variant has been associated with elevated IL-1β mRNA levels independently of inflammasome activation but not TLR ligands. These data suggest that defects in autophagy lead to increased IL-1β secretion and inflammation downstream of NOD2/CARD15.

In addition to activation via pathogen recognition receptors, type 1 and type II interferons are potent inducers
of the autophagy pathway. Interferon-γ can induce the expression of immunity-related guanosine triphosphate (IRG) proteins in mice while human IRG lack interferon response elements. IRGM is the most well-known member of the IRG family and is constitutively expressed at high levels in human cells. Similar to ATG16L1, genetic variants of IRGM have been significantly associated with CD risk; however, unlike ATG16L1, studies have been unable to correlate NOD2/CARD15 variants with IRGM.

Interestingly the lead SNP in IRGM (rs13361189) identified in CD is in perfect linkage disequilibrium with a 20 kb deletion upstream to the IRGM locus, hence the risk haplotype includes both deletion and SNP. Brest et al demonstrated that rs13361189 and the deletion are in perfect linkage disequilibrium with a synonymous exonic SNP (rs10065172) that alters IRGM levels and the binding of microRNA196. IRGM knockdown and microRNA196 overexpression prevents the targeting of adherent invasive E. coli to the autophagosome, suggesting that aberrant immune activity or loss of tolerance to commensal may explain the role of IRGM rs13361189 in CD.

While ATG16L1 is the most widely examined and well-characterized variant of the autophagy pathway associated with CD, recent studies have identified other important genes. Leucine-rich repeat kinase 2 (LRRK2) is a member of the leucine-rich repeat kinase family. Three studies have identified variants in LRRK2 associated with CD. The SNP, rs11175593, is located in a noncoding region on chromosome 12. Interestingly, reports have shown that LRRK2 may be a factor in autophagy via a scaffolding role during protein-protein interaction. ULK1 is a serine/threonine-protein kinase, which forms a complex with Atg1 and activates the Atg1/ULK1 complex via dephosphorylation of Atg1. This is the most upstream step of autophagosome formation. The SNP in ULK1 (rs12303764) was identified by Henckaerts et al in a population study of 1282 CD patients of Western European origin. NOD2/CARD15 variants were shown to have no impact on the association of rs12303764 with CD.

The identification of SNPs in LRRK2 and ULK1, and the already described role of ATG16L1, further establish the regulation of autophagy as an important cellular function in CD. However, more in-depth analysis of both LRRK2 and ULK1 SNPs must be performed to fully understand the functional consequence in CD pathogenesis.

Genetic predictors of Crohn’s disease clinical outcome
Since genetic factors do not change over time, the identification of genetic variants that may predict disease behavior and prognosis are superior to other predictive factors such as serological and clinical parameters. Despite the ever-growing number of susceptibility loci in CD, there are only a few variants associated with statistically significant clinical outcome and prognosis. In addition to its identification as a susceptibility locus, NOD2/D15 has also been identified in studies examining outcomes of CD. The presence of NOD2/
CARD15 has been associated with a more aggressive clinical course involving higher risk of intestinal strictures, earlier need for surgical intervention, and less postoperative disease-free intervals.\textsuperscript{13,73,79} NOD2/CARD15 was also found to be the most important factor for ileal location and stenosing and penetrating disease.\textsuperscript{72}

The presence of a genetic variant in Janus kinase 2 (JAK2) as a predictor for ileal involvement and stenosing disease was found in a cohort of 1528 European CD patients.\textsuperscript{72} Interestingly Prager et al.\textsuperscript{75} demonstrated that patients with the C risk allele within JAK2 rs10758669 have increased epithelial permeability, suggesting that the effect on disease behavior may be due to alterations in overall intestinal permeability.

Other genetic variants that impact disease behavior include an SNP in IRGM (rs4958847), which was significantly associated with frequency of surgery in a study of 66 patients with ileocecal CD, and a negative association was found between ileal disease and TLR1 S602I.\textsuperscript{79} Genetic variants may also impact the response to certain immune therapies; polymorphisms in multidrug resistant 1, TNF, and migration inhibitory factor genes have all been associated with sensitivity to corticosteroid,\textsuperscript{77,78} while variants in apoptosis genes have been used to predict the response to infliximab therapy.\textsuperscript{79} The identification of genetic variants could be an important tool in predicting disease behavior and responsiveness of certain patients to specific treatments; however, this may be complicated by the major role that other factors, such as the environment, plays in CD pathogenesis. It may be worthwhile to examine genetic variants alongside clinical, serological, and microbiological data in order to more accurately predict disease behavior.

Genetic variants in pediatric Crohn’s disease

Phenotypic differences between adult and pediatric onset CD suggest a different genetic predisposition. Pediatric IBD is characterized by extensive intestinal involvement and rapid early progression.\textsuperscript{80} The increased risk of CD among family members with early-onset disease and the stratification of disease by age of onset has been the focus of many studies in the effort to further ascertain the genetic influence of CD.

Genetic susceptibility may play a more important role in the etiology of early-onset IBD than in late-onset IBD, and therefore pediatric-onset IBD patients can be expected to have a higher frequency of gene mutations. De Ridder and colleagues\textsuperscript{81} investigated genetic polymorphism in CARD15 and DLG5 and described more frequent occurrence of polymorphisms, 3020insC, in CARD15 and SNP, rs3792876, in SLC22A4/5 in patients with pediatric-onset CD than in patients with adult-onset CD. Polymorphisms 3020insC in CARD15 and SNP rs2165047 in DLG5 were associated with specific phenotypes such as ileal and perianal disease, respectively. In another pediatric study in the GWAS literature,\textsuperscript{82} a previously unreported susceptibility locus for pediatric-onset IBD at 20q13 and 21q22 was identified, suggesting that TNFRSF6B is the most plausible candidate within the 20q13 locus, involved in both antigen-presenting cell differentiation and lymphocyte function. Autophagy-associated genes (ATG16L1 and IRGM) were also seen in early-onset CD cases, further confirming autophagy as an important contributor in the pathogenesis of pediatric CD.\textsuperscript{83}

Dubinsky et al.\textsuperscript{84} examined associations between the R381Q variant of the IL-23R gene and CD in a pediatric cohort and reported significant associations using a family-based study. Similarly, Van Limbergen et al.\textsuperscript{85} also reported associations with the same SNP in a Scottish cohort of pediatric CD using the case-control design. These reports provide strong evidence that the IL-23R gene is not only associated with adult-onset CD, but similar associations exist for children among most Caucasian populations.

Conclusions and perspectives

Currently there are more than 100 independent genetic loci contributing to the pathophysiology of CD. While many of these genes are considered risk factors, we are beginning to discover other genes that may help predict disease behavior as well as the response to certain immune therapies. Particularly interesting will be the further investigation of genes associated with pediatric early onset, thus allowing us the capability to identify children at a higher risk of disease and provide a more effective therapeutic approach. The last decade has witnessed a great expansion in our knowledge of the genetic variants in CD and the identification of many genetic factors that influence the innate immune response. These findings have provided an extensive framework for unraveling the genetic basis of disease susceptibility, clinical predictors of disease behavior, and biological processes of disease. As our understanding becomes more robust, greater insight into the genetic influence in CD will continue to be critical for the development of future therapeutic strategies.

Disclosure

The authors report no conflicts of interests in this work.

References


